



Image Retrieval with Reciprocal and shared Nearest Neighbors

Agni Delvinioti, Hervé Jégou, Laurent Amsaleg, Michael E. Houle

► To cite this version:

Agni Delvinioti, Hervé Jégou, Laurent Amsaleg, Michael E. Houle. Image Retrieval with Reciprocal and shared Nearest Neighbors. VISAPP–International Conference on Computer Vision Theory and Applications, Jan 2014, Barcelone, Portugal. hal-00907481

HAL Id: hal-00907481

<https://inria.hal.science/hal-00907481>

Submitted on 6 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Image retrieval with reciprocal and shared nearest neighbors

Agni Delvinioti^{†*}, Laurent Amsaleg[•], Hervé Jégou[†] and Michael Houle[∞]

[†]INRIA ^{*}University of Amsterdam [•]IRISA/CNRS [∞]NII, Tokyo

Keywords: Image search, reciprocal nearest neighbors, shared neighbors, image similarity

Abstract: Content-based image retrieval systems typically rely on a similarity metric between image vector representations, such as in bag-of-words or VLAD, to rank the database images in decreasing order of expected relevance to the query. However, the inherent asymmetry of a neighborhood based on k nearest neighbors negatively affects the retrieval accuracy, especially if the underlying representation of the vectors is highly-dimensional. This paper makes two contributions to alleviate this issue by better exploiting the neighborhood of each image. First, we revisit how to exploit the k -reciprocal nearest neighbors to produce a neighborhood that improves the one associated to the original metric. This strategy is simpler than concurrent prior works, yet it is both effective and less sensitive to parameters. Second, we propose to employ measures defined on sets of shared nearest neighbors in order to re-rank the shortlist. Both these methods are simple, yet they significantly improve the accuracy of image search engines on standard benchmarks.

1 INTRODUCTION

During the last decade, significant improvements have turned content based image retrieval systems from research laboratory prototypes into large scale, efficient and effective commercial products. In particular, the seminal “Video-google” paper (Sivic and Zisserman, 2003), by casting powerful local descriptors such as SIFT (Lowe, 2004) into a suitable vector representation, has made possible the use of numerous techniques previously introduced in text based information retrieval, such as inverted files or query expansion. Many systems are now based on this idea, possibly extended to cope with specific cases—these systems have proved to be powerful, allowing to quickly identify the images that are similar to a query, even when using databases comprising millions of images.

Various techniques try to enhance the result thanks to using geometry for re-ranking the shortlist (Fischler and Bolles, 1981; Jégou et al., 2008), by better defining the (larger) visual vocabulary (Nister and Stewenius, 2006), by altering the distance measure (Jégou et al., 2007), by aggregating local features in clever ways (Perronnin et al., 2010; Jégou et al., 2010), etc. Such approaches do improve the quality of the results, yet, there is still room for improvement.

One fundamental reason undermining this quality problem comes from the asymmetry of the k -

nearest neighbor (k -nn) criterion used to identify similar items. Indeed, it is very possible that if a is the nearest-neighbor of b , b might not be the nearest neighbor of a . This trivially extends to the case where k neighbors are considered. Overall, because of this asymmetry, a traditional k -nn neighborhood is likely to contain noisy data, i.e., points that are quite dissimilar. These false positives, in turn, severely lower result quality.

Several approaches try to compensate this asymmetry by augmenting the k -nn criterion with information derived from studying the neighborhood of points. For example, Qin *et al.* in (Qin et al., 2011) define neighborhoods according to the notion of k -reciprocal nearest neighbors instead of simply using the primary cosine similarity. In another context, Houle *et al.* in (Houle et al., 2010) propose to use the notion of shared nearest neighbors. These two approaches are effective in improving the quality of similarity retrievals.

This paper follows that route and proposes two contributions aiming at compensating for the fundamental asymmetry of the k -nn measure. The first contribution is the definition of three robust and stable extended similarity measures for comparing the neighborhoods of the candidates in the shortlist, for then re-ranking them. The second contribution is the definition of a maximum reciprocal rank criterion to construct a better shortlist containing more highly relevant images. When used either in isolation or in a combined manner, these two contributions signifi-

cantly improve the accuracy of image search engines when evaluated against traditional benchmarks.

The paper is structured as follows. Section 2 gives an overview of the work that is most related to this paper. Section 3 details our two contributions. Then, Section 4 gives experimental evidence of the improvements observed when using our techniques on standard benchmarks. Section 5 concludes the paper.

2 RELATED WORK

Qin *et al.* in (Qin et al., 2011) analyze the k -reciprocal nearest neighbor structure in the image space by treating different parts of the shortlist with different distance measures. Given a query image q , they separate the database into two disjoint datasets, the *close-set* which contains images highly related to the query and the *far-set* simply denoting the rest of the database. The *close-set* of q is gradually constructed in an expanding manner. Two adjacency criteria between q and its k -reciprocal neighbors allow to decide to add or not images to the *close-set*: only images which are connected to at least half of the *close-set* can bring in their neighbors—this is emphasizing high adjacency; Another relaxed condition allows to bring in the neighbors of less strongly connected images—this is emphasizing low divergence among neighborhoods. The *close-set* is used to re-rank images from the *far-set* according to the connectivity degree of *far-set* images to *close-set* ones.

This approach has various parameters that are difficult to tune—we were unable to exactly reproduce the results from (Qin et al., 2011); Furthermore, the expansion process for the *close-set* is quite costly. While the *close-set* nicely makes use of highly related neighbors, the *far-set* suffers from the asymmetric primary cosine similarity ranking biased by an arbitrary cutoff value. Last, the dependency on the reciprocal neighborhood size k is significant, resulting in an unstable performance behavior as k increases. Overall, acquiring a decoupled strategy considering different parts of the database to exploit structural information over neighborhoods turns out to be insufficient.

Houle *et al.* in (Houle et al., 2010) introduces shared nearest neighbor approach as a secondary rank-based similarity measure. It provides a more stable and robust metric than traditional noisy measures, especially in high dimensional spaces. Here, the similarity value of an object pair is a function of the number of data objects in the common intersection of fixed-sized neighborhoods, determined by a primary conventional similarity measure. That primary similarity measure can be any function provid-

ing a ranking of the data objects relative to the query, typically a cosine similarity.

The *relevant-set correlation* RSC clustering model (Houle, 2008) adopts such a shared-neighbor scheme to account for well-associated items in the grouping procedure. (Hamzaoui et al., 2013) also builds on shared-neighbors. They designed a bipartite shared-neighbors clustering algorithm for suggesting additional highly relevant object-based visual queries improving the search experience.

All three approaches sketched here consider crucial to take into account the notions of shared neighborhood. They all try to compensate for the difficulties induced by high dimensions with an enhanced but simple way to extract adjacency and structural information among neighborhoods in the image space or among object seeds in the object space. They all “denoise” the primary traditional similarity measure, in a second processing step, in order to provide more reliable matches.

3 OUR APPROACH

Overall, two contributions are proposed in this paper. They both try to compensate for the asymmetry of the k -nn measure. Each can be used in isolation, or in a combined manner. The first contribution builds on the contents of the shortlist containing the k images that have been found the most similar to a query. It consists in a specific re-ranking mechanism that compares the neighborhoods of the images that are in the shortlist. This mechanism integrates, for increasing values of k , the knowledge acquired from determining the shared nearest neighbors from the images in the shortlist. This knowledge somehow corresponds to the structural image similarity of the entire database. This mechanism is presented in Section 3.1, where we detail three neighborhood comparison metrics.

The second contribution, presented in Section 3.2, aims at producing a higher quality shortlist to start from. It consists in the definition of a maximum reciprocal rank criterion to construct the shortlist containing the points that are in the neighborhood of a query, taking into account pair-wise information.

3.1 Comparing neighborhoods

In traditional systems, the k -nn of a query q are determined directly from the cosine similarity; Then, ranks are assigned to the candidates in the shortlist accordingly. We define this list as being $\mathcal{N}_k(q)$. It has been reported that this shortlist often contains irrelevant

images and/or quite similar images poorly ranked, in part due to the asymmetry of the k -nn criterion.

It is possible to re-rank the images in the shortlist by considering the number of similar images that are shared by the members of this shortlist. This idea refers to the notion of shared nearest neighbors proposed by M. Houle *et al.* (Houle et al., 2010).

3.1.1 Shared nearest neighbors

Once $\mathcal{N}_k(q)$ determined, it is then possible to parse that shortlist and, for each element, determine the overlap degree of their fixed-size neighborhood. Given two images t, u in the shortlist of q , their neighborhoods' overlap is defined as the cardinality of the intersection of the images that are in their respective neighborhoods. The intersection is defined as:

$$SNN_k(t, u) = \mathcal{N}_k(t) \cap \mathcal{N}_k(u) \quad (1)$$

and the cardinality is thus simply $|SNN_k(t, u)|$.

3.1.2 Metrics for neighborhoods

This pairwise knowledge for all the images in the shortlist can then typically be used within a similarity measure for re-ranking that shortlist. Two images in the shortlist that share a lot of similar images from the database are likely to be more similar than two database images little sharing. Comparing the neighborhoods of the images in the shortlist can therefore be mapped to comparing sets of images. In this paper we use three such metrics:

- *Jaccard*: A traditional metric to measure the similarity between sets is the Jaccard coefficient:

$$j_k(x, y) = \frac{|SNN_k(x, y)|}{|\mathcal{N}_k(x) \cup \mathcal{N}_k(y)|} \quad (2)$$

It ranges from 0 to 1, with $j_k(x, y) = 1$ implying that the images share exactly the same set of neighbors, in which case we assume they are highly likely to be exactly similar. By taking into account the union of the sets, j_k measures the distribution of the shared neighbors.

- *Set Correlation*: Another possible metric is the set correlation measure defined by Houle (Houle, 2008). This measure is, for x, y in the database \mathcal{D} .

$$sc_k(x, y) = \frac{|\mathcal{D}|}{|\mathcal{D}| - k} \left(\frac{|SNN_k(x, y)|}{k} - \frac{k}{|\mathcal{D}|} \right) \quad (3)$$

Neighbors appearing in both $\mathcal{N}_k(x)$ and $\mathcal{N}_k(y)$ support the correlation. In contrast, neighbors joining only one of the neighborhood and not the other detract from the correlation. Compared to j_k , the set correlation metric does not take the union of the sets into account.

- *Sigmoid*: The two measures previously defined fail to differentiate strong similarities observed when k is small from weaker similarities when k is large. In this later case, neighborhoods typically include a significant proportion of irrelevant neighbors and thus are less informative and reliable than when k is small. By construction, j_k and sc_k increase with k , and fail to reflect that the associated neighborhoods become noisy.

A sigmoid type-of function can be used for comparing sets and at the same time differentiate strong from weak similarities, mitigating the influence of (large) k . The function we define is:

$$sgm_k(x, y) = \frac{1}{1 + \exp(-a * (\frac{|SNN_k(x, y)|}{k} - b))}, \quad (4)$$

$\frac{|SNN_k(x, y)|}{k}$ is the normalized intersection based on neighborhood size k . a is the slope of the curve. $b = \exp(-\frac{k}{n})$ is a decreasing function of k , avoiding being biased when k gets large.

The value of a controls how are mapped the k similarities. When $a < 1$, then the resulting mapping is very gradual. In contrast, higher values for a create a much sharper mapping as the slope becomes more steep. The extreme case is when the sigmoid function is turned to a step function. Since our goal is to benefit on the one hand from the profitable behavior of Jaccard and Set Correlation for small k similarity values and on the other hand from the thresholding of similarities when k is large, then the natural choice is to have $a = 1$.

3.1.3 Extending neighborhoods metrics

The motivation for proposing the sigmoid metric comes from the desire to care about the influence of the poorly relevant images (having their ranks closer to k) eventually dominating the influence of the highly relevant images (having their ranks closer to 1). The Jaccard and the Set Correlation metrics are oblivious to the rank in the shortlist at which neighborhoods starts to diverge; it is only concerned with the value of k . Note that, however, the point of inflection in the sigmoid function allows to change the relative influence of the neighborhood, yet, it is very sensitive to the value of k .

We therefore define a more robust process for computing the similarity of the neighborhoods obtained from analyzing the shortlist. This process retains high quality/reliable votes that are likely near the top of the shortlist while it also accounts for diverging neighborhoods closer to the bottom of the list. The process refines the shortlist accounting for varying neighborhood sizes. It tries to use all the con-

tributions from the observed sharings in a balanced, unified scheme, in order to extract a similarity score.

To implement this process, we propose three extended metrics for comparing the neighborhoods. The idea is to integrate in a single measure the neighborhood that can be observed within an increasing horizon. We therefore compare the sets of similar images for each element of the shortlist varying k' such that $1 \leq k' \leq k$. The extended Jaccard, extended Set Correlation and extended Sigmoid measures averaging the similarity values observed within an horizon that is increasing step by step, until it reaches k are thus, respectively:

- *Extended Jaccard:*

$$\bar{j}_k(x, y) = \sum_{k'=1}^k j_{k'}(x, y) \frac{1}{\sum_{l=1}^{k'} \delta_l(x, y)}, \quad (5)$$

with

$$\delta_l(x, y) = \begin{cases} 1, & \text{if } |SNN_l(x, y)| > 0 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

- *Extended Set Correlation:*

$$\bar{sc}_k(x, y) = \sum_{k'=1}^k \frac{sc_{k'}(x, y)}{k'} \quad (7)$$

- *Extended Sigmoid:*

$$\bar{sgm}_k(x, y) = \sum_{k'=1}^k \frac{sgm_{k'}(x, y)}{k'} \quad (8)$$

It is possible to slightly alter the definitions of the above measures by changing the starting value for k' , i.e., $k_0 \leq k' \leq k$. k_0 defines the starting point for the similarity averaging process. The measures above are with $k_0 = 1$. k_0 has an obvious mathematical influence when averaging. k_0 has also a somehow higher level influence as it captures one intimate characteristics of each dataset: The number of truly matching elements. When there are quite a lot of truly matching elements per query (this *low diversity* case can be observed for the Oxford5k and Paris6k datasets with hundreds matches per query, see Section 4.1), then it matters to start right away with a large k' as this increases the average score, computed with fewer iterations. In contrast, when there are only few true matches (this *high diversity* case can be observed for the Holidays dataset, see Section 4.1), then setting k_0 to a small value allows to take these matches into account.

3.2 Maximum reciprocal rank

It has often been reported that $\mathcal{N}_k(q)$ contains some irrelevant images because the k -nn is by nature asymmetric. To address that issue, Qin *et al.* in (Qin et al.,

2011) propose a solution built on top of the primary cosine similarity measure. They use the notions of *close-set*, *far-set* and reciprocal neighbors to determine a higher-quality shortlist for q .

In contrast to their approach which is somehow costly and quite unstable, we propose here another way to use the notion of reciprocity for determining a highly relevant neighborhood for q .

We first define $\text{rank}_x(y)$ as the rank of the image y when the database is queried by x using a traditional cosine similarity criterion. Conversely, $\text{rank}_y(x)$ is the rank of x when y is the query. $\text{rank}_x(y)$ is typically termed *forward rank* while $\text{rank}_y(x)$ is termed *backward rank*. We can then define $r(\cdot, \cdot)$, a dissimilarity measure, indicating the bidirectional relationship implied by reciprocity:

$$r(x, y) = \max_{x, y \in \mathcal{D}} (\text{rank}_x(y), \text{rank}_y(x)), \quad (9)$$

It is now possible to define $\mathcal{R}_k(x)$, the k Maximum Reciprocal Rank of an image $x \in \mathcal{D}$ as:

$$\mathcal{R}_k(x) = k\text{-arg min}_{y \in \mathcal{D}} r(x, y), \quad (10)$$

$\mathcal{R}_k(x)$ allows to determine what are the images that it is valuable to consider in the vicinity of x : these images are not only reciprocal neighbors of x , but the extent of reciprocity is strictly bounded by the value of k . This reciprocity-based neighborhood is a much stronger indication of similarity than what the asymmetric k -nn primary similarity measure directly determines.

For these reasons, we propose to use $\mathcal{R}_k(q)$ instead of $\mathcal{N}_k(q)$ for determining the contents of the shortlist corresponding to the query q .

3.3 Discussion

The extended measures that were defined in Section 3.1.3 rely on two different notions of neighborhoods. The first notion corresponds to the *direct* neighborhood of the query that is used to determine the shortlist of similar images. The second notion corresponds to the neighborhood of the images inside that shortlist for building on their own shared neighbors. This is the *indirect* neighborhood of the query.

The criterion for determining the neighborhood of one image (whether this image is the query and/or one image from the shortlist) can be either based on the traditional k -nn scheme or based on the Maximum reciprocal rank scheme defined above in Section 3.2. In this later case, neighborhoods are likely to be relevant compared to the asymmetric k -nn case.

It should therefore be clear that these two mechanisms—determining neighborhoods and ex-

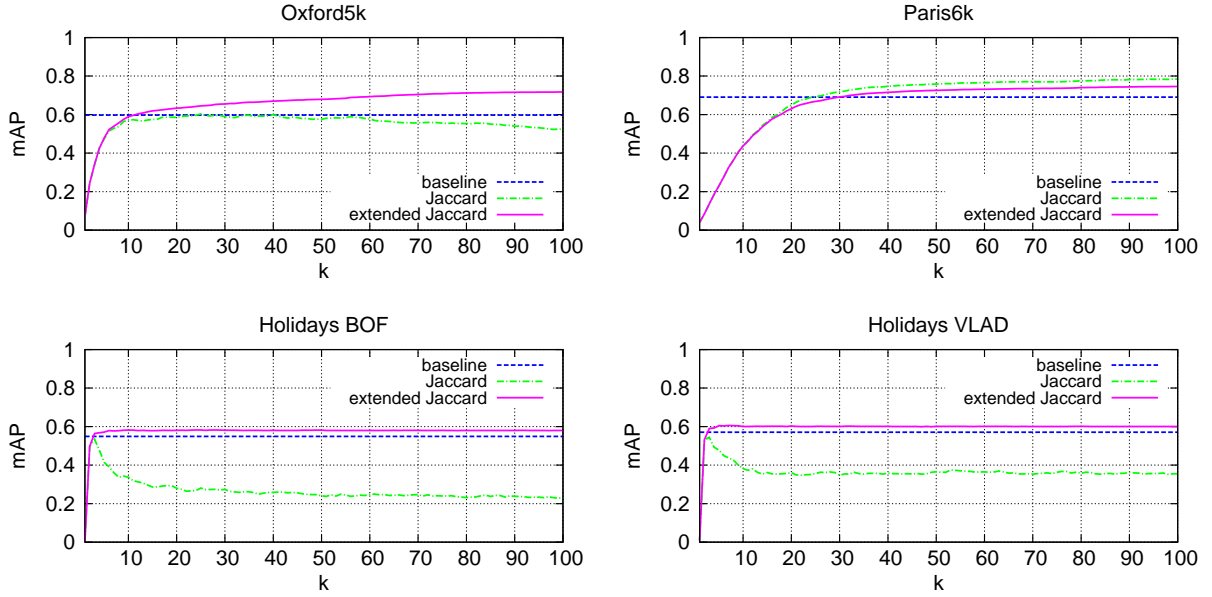


Figure 1: Comparing the mAP determined from using the baseline (cosine+k-nn) with the mAP obtained using SNN_k relying on \mathcal{N}_k together with the Jaccard j_k or the extended Jaccard \bar{j}_k measures. Comparison performed against the Oxford5k, the Paris6k and the BOF and VLAD versions of the Holidays image sets, varying k .

exploiting neighborhoods—can be used either independently one from the other, or in a combined manner. For example, it is possible to use \bar{j}_k with $\mathcal{N}_k(q)$, or to use \bar{j}_k in combination with using $\mathcal{R}_k(q)$ as well. As this will be described in the experimental section, it is possible to observe that the quality increases when using one of the extended neighborhood comparison measure to then re-rank. When determining neighborhoods using the maximum reciprocal rank and then applying our re-ranking process, then further improvements can be observed. This is demonstrated by the experiments presented next.

4 EXPERIMENTS

In this section we report the performance of the approach presented above. We start by reporting the improvements on the quality of the retrievals when using the extended metrics comparing shared neighborhoods. We then compare the performance of these metrics when building the shortlist using $\mathcal{R}_k(q)$ instead of $\mathcal{N}_k(q)$, that is, when relying on the maximum reciprocal rank criterion rather than the asymmetric traditional k -nn. Then, we show the impact of setting $k_0 > 1$. First, however, we describe the datasets used in the experiments, the images features extracted as well as the state-of-the art baseline against which all comparisons are done.

4.1 Experimental setup

We use the three following popular datasets for our experiments. They have also been used in many other publications and they are all available online, facilitating the reproducibility of results.

Holidays (1491 images, 500 queries). This dataset contains high resolution, personal holiday photos provided by INRIA (Jégou et al., 2008) and it is divided in small groups of images showing the same object or the same scene. Each query is compared to the other 1490 images in a leave-one-out fashion.

Oxford5k (5062 images, 55 queries). It consists of images corresponding to 11 distinct buildings in Oxford (Philbin et al., 2007). All queries are given accompanied by specific bounding boxes and they are compared to all the database images.

Paris6k (6412 images, 55 queries). These images come from Flickr through a search procedure focusing on particular landmarks in Paris (Philbin et al., 2008). It has also bounding boxes and queries are compared to all images.

SIFT descriptors were extracted from the three image sets using a Hessian-Affine detector. Descriptors are then clustered according to a visual vocabulary, specified below for each image set.

The high dimensional features describing the Holidays image set were assigned to a vocabulary comprising 200k visual words that was found online,

Table 1: Comparing the mAP determined from using the baseline (cosine+ k -nn) with the mAP obtained relying on \mathcal{N}_k together with extended Jaccard \bar{j}_k , Set Correlation \bar{sc}_k and Sigmoid \bar{sgm}_k measures. $k = 100$ on Oxford5k and Holidays, BOF.

Method	Oxford5k	Holidays
baseline	0.598	0.549
\mathcal{N}_k & \bar{j}_k	0.699	0.580
\mathcal{N}_k & \bar{sc}_k	0.697	0.580
\mathcal{N}_k & \bar{sgm}_k	0.722	0.587

see (Jégou et al., 2008). All the experiments presented below that use this bag-of-features representation of the Holidays image set are labeled “Holidays BOF”.

Many systems have tried using a more sophisticated representation of the local features extracted from an image set. Typically, aggregating local features gives better result quality. We thus felt necessary to apply our approach to these sophisticated description schemes and evaluate the gains. We therefore computed VLAD features over the Holidays image set (Jégou et al., 2010), clustering them to 64 centroids. The experiments using the VLAD representation of Holidays are labeled “Holidays VLAD”. The other reason for using VLAD is to determine the ability of our technique to also improve quality when using lower dimensional features.

The features computed from the Oxford5k and Paris6k image sets were assigned to two different visual vocabularies that were kindly given to us by Qin *et al.*. The Oxford5k feature set was assigned to 1M visual words, while the Paris6k feature set was assigned to 500k visual words.

The baseline for evaluating our approach is obtained by computing the mean average precision (mAP) on the results returned when probing the databases with the queries. The search process is here a simple k -nn search based on the cosine similarity. At search time, L_2 normalization is applied to the cosine similarity and a traditional *tf-idf* weighting scheme is used. We typically ran many series of retrievals varying k . Note that we do not use any post-processing step checking the geometrical consistency of matches.

4.2 Jaccard vs. extended Jaccard

We start showing the performance of our extended neighborhood metrics by comparing the results obtained when using on the one hand SNN_k relying on $\mathcal{N}_k(q)$ together with the Jaccard coefficient j_k and on the other hand SNN_k relying on $\mathcal{N}_k(q)$ together with

its extended version \bar{j}_k . We start with Jaccard as it is a very traditional coefficient for measuring the similarity of sets. Figure 1 shows the mAP for the baseline, the Jaccard and the extended Jaccard over the Oxford5k, the Paris6k and the two versions of the Holidays (BOF and VLAD) datasets, varying k . $k_0 = 1$.

Compared to the baseline which is simply a cosine with k -nn and *tf-idf*, the j_k Jaccard metrics return better results on the Paris6k set, as the mAP increase shows. This is due to the use of the shared neighborhood SNN_k detailed Section 3.1.1 for re-ranking the shortlist $\mathcal{N}_k(q)$. In contrast, performance are worse than the baseline and degrade as the value of k increase for Oxford5k and for the two versions of the Holidays dataset. A careful analysis of the experimental logs and the image sets shows that SNN_k enhances the results when the databases contain images that are not that diverse. This is the case for Paris6k, as it includes a limited number of landmarks. This is not at all the case for Holidays which contains extremely diverse images. With Holidays, only very few images are indeed similar to each query, and these images are in general very well ranked. Therefore, as k increases, the shortlist contains more and more unrelated matches, adding noisy matches to SNN_k . This behavior is clearly visible on the curves showing the performance of the Jaccard coefficient where the mAP sharply decreases for very small k . Note also that this behavior can be observed on the Oxford5k dataset, less pronounced, however, explaining why the performance degrade as k increases.

We now turn to the extended version of the Jaccard coefficient, \bar{j}_k . The extension, integrating various neighborhoods observed within an increasing horizon, exhibit a much better behavior than the basic Jaccard. \bar{j}_k is able to preserve the reliable matches originating from the highly relevant shared neighborhoods observed from the strong similarities of the elements in the shortlist when k is small, thus providing a smooth and stable behavior. Note that no improvement can be observed when using the Paris6k image set—This is again caused by the very low diversity of the images and because many images indeed match any query.

This first experiment shows the extended version of Jaccard not only outperforms but has also a more desirable behavior than the one of the traditional Jaccard. We also evaluated the performance of sc_k versus \bar{sc}_k as well as sgm_k versus \bar{sgm}_k —the corresponding plots are not shown, however. Overall, the extended versions of the three neighborhood comparison metrics always outperform their non-extended counterpart and always show a smoother behavior. For these reasons, we will always compare in the remainder of

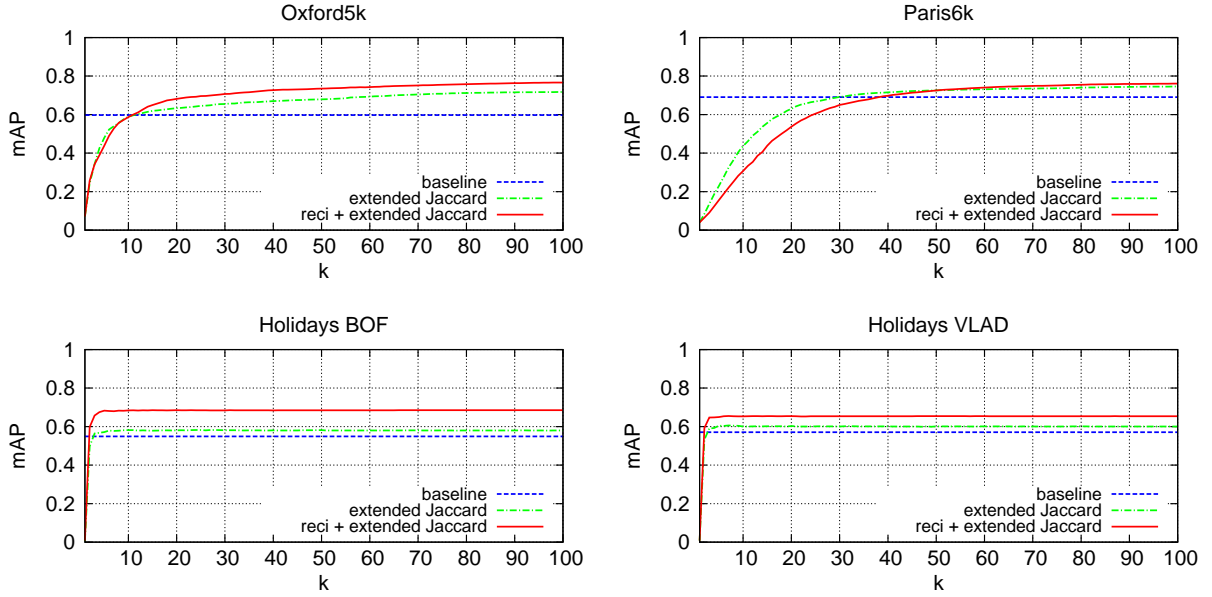


Figure 2: Comparing the behavior of \bar{j}_k when the shortlist for queries is based on \mathcal{N}_k or \mathcal{R}_k . Comparison performed against the Oxford5k, the Paris6k and the BOF and VLAD versions of the Holidays image sets, varying k .

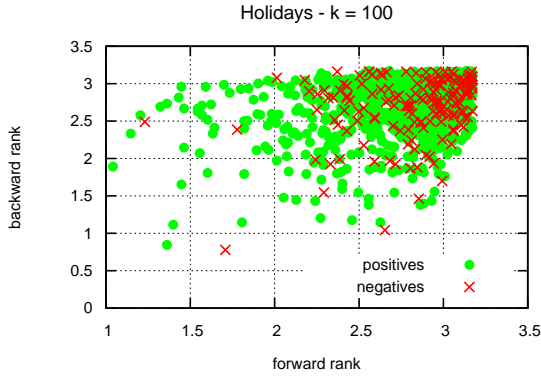


Figure 3: Illustrating the maximum reciprocal rank. Holidays image set, BOF features. $k = 100$ for all 500 queries. “positives” (green circles) and “negatives” (red crosses) are determined from the groundtruth. Log-scales.

this paper the performance of the baseline to the ones of the extended versions of the Jaccard, the Set Correlation and the sigmoid metrics, \bar{j}_k , \bar{sc}_k and \bar{sgm}_k respectively.

4.3 Comparing extended metrics

The second experiment presented here compares the performance of the three extended metrics against the baseline when $k = 100$. Overall, Table 1 shows that \bar{j}_k , \bar{sc}_k and \bar{sgm}_k outperform the baseline without suffering from the noisy data even when using a large value for k . This table shows only the results for the Oxford5k and the Holidays image sets; The perfor-

mance for the Paris6k is quite similar as the one for Oxford5k, this is due to low image diversity, and the performance for the VLAD version of Holidays is almost identical to the BOF version. Note that \bar{sgm}_k shows the best performance here.

4.4 Using reciprocal neighbors

The third experiment presented in this section compares the mAP improvements when the shortlist to start from is constructed using the cosine similarity (this is \mathcal{N}_k), or using the maximum reciprocal rank criterion (this is \mathcal{R}_k). Then, the elements of one or the other shortlist are re-ranked according to the extended metrics presented and evaluated above.

Consistent with the observations made earlier, the three extended metrics behave quite identically, regardless of the method used first to construct shortlist—consequently, we discuss in this third experiment only j_k and \bar{j}_k , conclusions apply to the two other metrics and their extended counterpart.

Figure 2 plots the mAP for the all the datasets we used here. Overall, it shows the performance of \bar{j}_k using the maximum reciprocal rank criterion \mathcal{R}_k (indicated by “reci+extended Jaccard”) clearly outperforms \bar{j}_k when it otherwise uses the traditional k -nn criterion for building the shortlist \mathcal{N}_k (indicated by “extended Jaccard”). The gain is extremely significant on the Oxford5k image set, known to be quite challenging.

Reciprocity, but when used according to the maximum rank rule, clearly allows the search to build a

Table 2: mAP observed for various search strategies and re-ranking techniques; Holidays, Oxford5k and Paris6k sets. $k_0 = 1$.

Method	Oxford5k	Paris6k	Holidays	
	BOF	BOF	BOF	VLAD
baseline	0.598	0.691	0.549	0.571
Qin <i>et al.</i> (Qin et al., 2011)	0.814	0.803	-	-
\mathcal{N}_k & \bar{j}_k	0.701	0.752	0.582	0.606
\mathcal{N}_k & \overline{sc}_k	0.700	0.748	0.581	0.602
\mathcal{N}_k & \overline{sgm}_k	0.724	0.783	0.589	0.607
\mathcal{R}_k & \bar{j}_k	0.737	0.768	0.685	0.655
\mathcal{R}_k & \overline{sc}_k	0.734	0.765	0.684	0.654
\mathcal{R}_k & \overline{sgm}_k	0.746	0.804	0.687	0.660

much higher quality shortlist to start from. Then re-ranking does a good job as the material it starts with is highly relevant. Overall, reciprocity and the extended neighborhood comparison give strong indications that it is possible to create a balanced combination in a unified scheme of a global and structural knowledge from the analysis of the sharing of neighbors with more local pairwise information given by reciprocity.

The behavior of the maximum reciprocal rank rule is illustrated by Figure 3. Here, we plot the *forward* and *backward* ranks for the 100 nearest neighbors determined for all 500 queries. Positives green circles materialize the images truly relevant to queries, according to the groundtruth. Negative red crosses materialize images erroneously considered as relevant when checked against the groundtruth. Note the log-scales. The density of positives is high close to the diagonal; most negatives have high *forward* and *backward* ranks.

4.5 Varying initial neighborhood size

This last experiment shows the impact of the initial neighborhood size k_0 used when computing the extended metrics. It was $k_0 = 1$ in all previous experiments. Figure 4 shows the influence of k_0 on the mAP when using \mathcal{N}_k (the two top graphs) or \mathcal{R}_k (the four bottom graphs) with the Paris6k and the Oxford5k image sets. The k_0 values for each image set used in this Figure are the ones providing the most indicative results. In all cases, the quality improvements are significant. It compensates particularly well the (poor) asymmetric neighborhood defined with \mathcal{N}_k .

4.6 Summary of experiments

Table 2 summarizes the results obtained when running all our evaluations, a subset of them were already

Table 3: mAP observed for \mathcal{R}_k and the extended re-ranking techniques considering a varying initial neighborhood size k_0 on Oxford5k and Paris6k image sets.

Oxford5k		
Method	$k_0 = 1$	$k_0 = 20$
\mathcal{R}_k & \bar{j}_k	0.737	0.779
\mathcal{R}_k & \overline{sc}_k	0.734	0.777
\mathcal{R}_k & \overline{sgm}_k	0.746	0.761
Paris6k		
Method	$k_0 = 1$	$k_0 = 80$
\mathcal{R}_k & \bar{j}_k	0.768	0.820
\mathcal{R}_k & \overline{sc}_k	0.765	0.820
\mathcal{R}_k & \overline{sgm}_k	0.804	0.812

presented and discussed above. This table shows the best mAP values that were observed when running our techniques for a k that ranges between 1 and 200. This table also compares the approaches we propose to the one designed by Qin *et al.* (Qin et al., 2011). Note, however, that our results slightly differ from the ones presented by Qin *et al.* in their paper when using the Oxford5k and the Paris6k image sets—despite our efforts, we could not *exactly* reproduce their work, possibly in part because the visual vocabularies provided on-line might not be the one used in their publication. Note that here $k_0 = 1$. The best mAP found overall are in bold font.

Table 3 reports the results obtained when setting k_0 to specific values larger than 1 and when using \mathcal{R}_k . Results are only for Oxford5k and Paris6k, where setting k_0 significantly boosts the mAP. This is not the case for Holidays due to its high diversity (its small number of truly matching elements per query). Experiments ran with \mathcal{N}_k show the mAP improves for Paris6k when $k_0 = 60$ and goes to 0.790 for \bar{j}_k , to

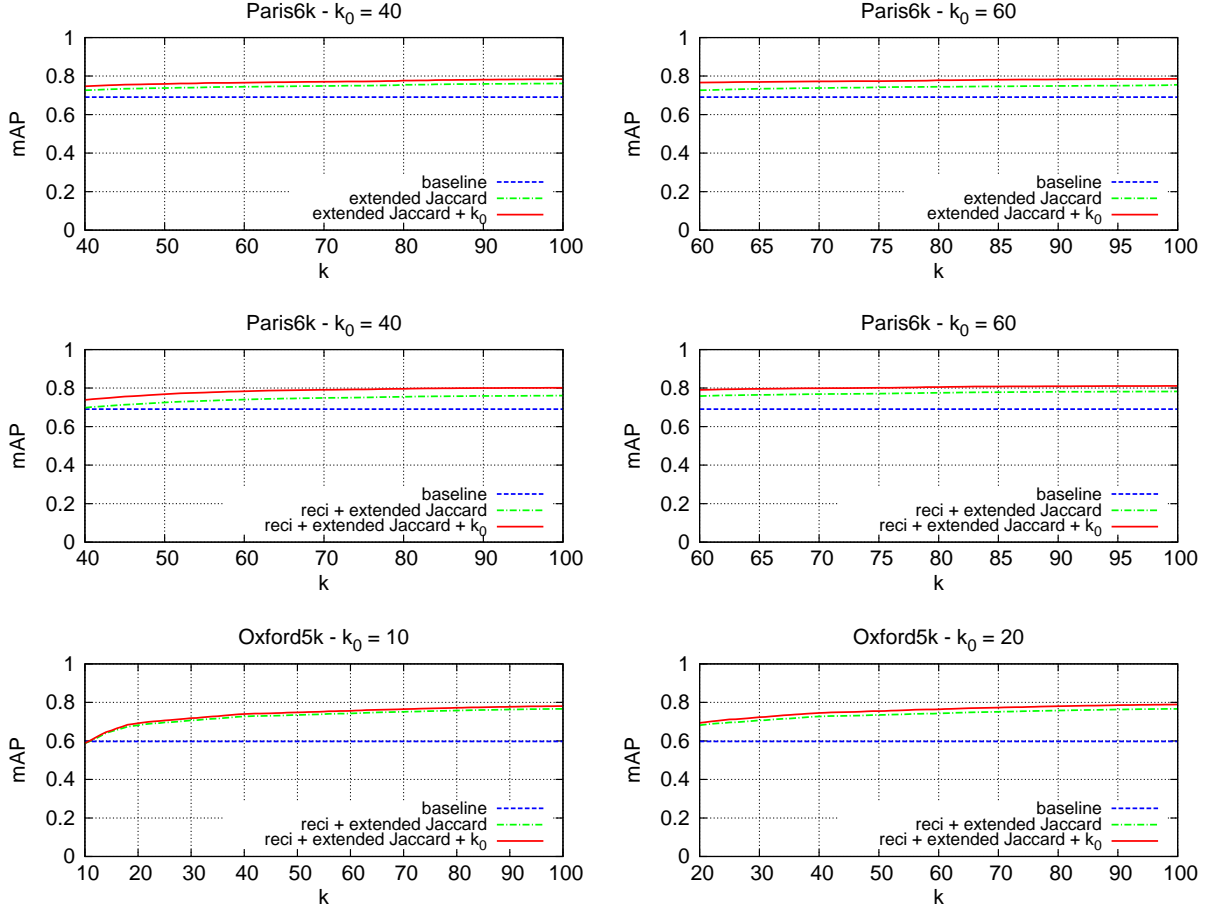


Figure 4: Illustrating the impact of the initial neighborhood size k_0 on the behavior of \bar{j}_k when the shortlist for queries is based on \mathcal{N}_k or \mathcal{R}_k . $k_0 = \{40, 60\}$ for Paris6k and $k_0 = \{10, 20\}$ for Oxford5k.

0.789 for \overline{sc}_k and to 0.781 for \overline{sgm}_k . All the results discussed in this paragraph are the best we observed when running experiments with $k \leq 200$.

4.7 Lessons

Three lessons can be drawn from our work:

1. *Shared Nearest Neighbors are useful.* It is clear from the experiments and from the tables that taking into account the neighbors shared by the images in the shortlist dramatically improves quality. Spectacular improvements are with the Paris6k image set.
2. Extending the metrics used to compare the neighborhoods at re-ranking time boosts the mAP. Furthermore, it avoids, to a large extent, hurting the performance when the value of k increases.
3. Building the shortlist using the maximum reciprocal rank clearly outperforms the other traditional methods based on the asymmetric k -nn criterion.

Figure 5 illustrates the visual improvements by showing the top 5 images returned when querying the Oxford5k, the Paris6k and the Holidays image sets with one specific query. In this Figure, and for each dataset, the top row shows what the system returns when it uses \mathcal{R}_k & \overline{sgm}_k on Holidays \mathcal{R}_k & \bar{j}_k on Oxford5k and Paris6k while the bottom row shows what is returned when using the simple cosine similarity over the BOF model together with the asymmetric k -nn paradigm. It is clear that our method avoids, in most cases, false positives.

5 CONCLUSION

This paper presents three extended metrics for comparing the neighborhood of the images that are in the shortlist of images similar to a query. That comparison takes into account an increasing horizon for assessing the sharing of neighbors. This results in a re-ranking of the images in the shortlist that significantly

improves the mAP observed over state-of-the-art standard image benchmarks. This paper also presents a maximum reciprocal rank criterion allowing to create shortlists containing highly relevant images. Both techniques, used in isolation or in a combined manner, outperform standard techniques.

Overall, compared to the work presented in (Qin et al., 2011), our approach provides a quite simple and uniform framework for merging the structural information from studying the neighborhood of similar images with their similarity to the query point. Furthermore, our re-ranking procedure remains free of any complicated parameter tuning (k_0 can be ignored or set to a medium value by default) and does not include any optimization process, keeping its complexity low. We need, however, to precompute and store the maximum reciprocal ranked lists. The memory overhead therefore grows linearly with the database size, exactly as for (Qin et al., 2011).

ACKNOWLEDGEMENTS

This work was done partly in the context of the ANR project FIRE-ID (ANR-12-CORD-016).

REFERENCES

- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6).
- Hamzaoui, A., Letessier, P., Joly, A., Buisson, O., and Boujemaa, N. (2013). Object-based visual query suggestion. *Multimedia Tools and Applications*.
- Houle, M. E. (2008). The relevant-set correlation model for data clustering. *Statistical Analysis and Data Mining*, 1(3).
- Houle, M. E., Kriegel, H. P., Kröger, P., Schubert, E., and Zimek, A. (2010). Can shared-neighbor distances defeat the curse of dimensionality? In *Scientific and Statistical Database Management*.
- Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*.
- Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *CVPR*.
- Jégou, H., Harzallah, H., and Schmid, C. (2007). A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2).
- Nister, D. and Stewenius, H. (2006). Scalable recognition with a vocabulary tree. In *CVPR*.
- Perronnin, F., Liu, Y., Sánchez, J., and Poirier, H. (2010). Large-scale image retrieval with compressed fisher vectors. In *CVPR*.
- Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *CVPR*.
- Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*.
- Qin, D., Gammeter, S., Bossard, L., Quack, T., and Gool, L. V. (2011). Hello neighbor: accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR*.
- Sivic, J. and Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *ICCV*.

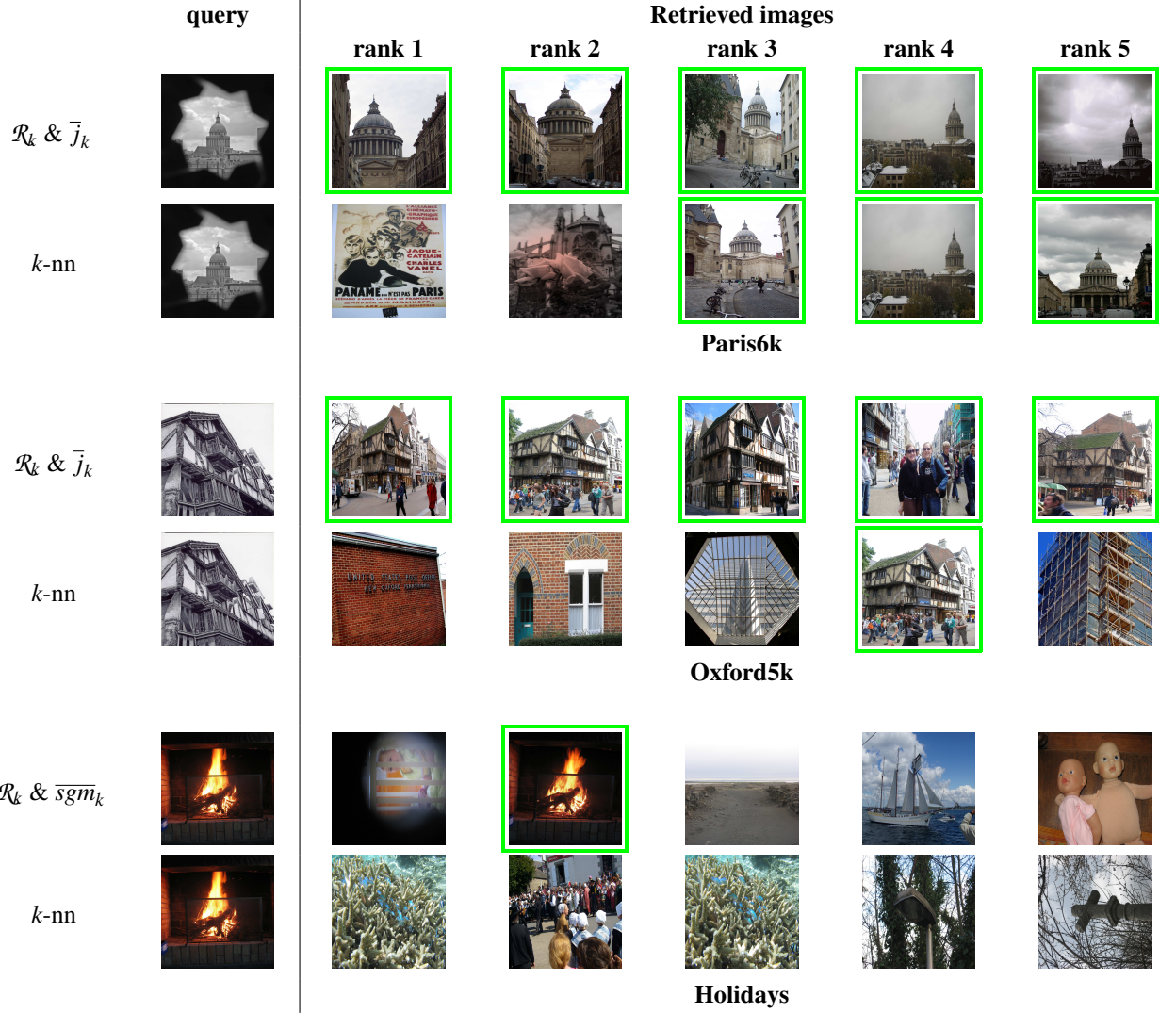


Figure 5: Retrieval examples considering one query per dataset (Oxford5k, Paris6k and Holidays) presented the difficulty characterizing each dataset, from high to low one. The ranking lists are obtained using our method regarding extended Jaccard measure for Oxford5k and Paris6k and extended Sigmoid measure for Holidays as the ones inducing the superior performance results. We can observe that our method improves significantly ranking while provides reliable matches even when simple k -nn scheme fails to discover any matches.